



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 1/21, 9/90, 15/61, 15/63</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 99/27074</b> <b>(43) International Publication Date:</b> 3 June 1999 (03.06.99)
<b>(21) International Application Number:</b> PCT/US98/24805 <b>(22) International Filing Date:</b> 20 November 1998 (20.11.98)  <b>(30) Priority Data:</b> 60/082,418 20 November 1997 (20.11.97) US  <b>(71) Applicant:</b> SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US).	<b>(72) Inventors:</b> ZALACAIN, Magdalena; 126 Woodmint Drive, West Chester, PA 19380 (US). BROWN, James, R.; 9 Robins Lane, Berwyn, PA 19312 (US). BISWAS, Sanjoy; Apartment B4, 77 S. Valley Road, Paoli, PA 19301 (US). THROUP, John, P.; 506 Brooke Drive, Royersford, PA 19468 (US). LAWLOR, Elizabeth, J.; 260 Lapp Road, Malvern, PA 19355 (US). MOONEY, Jeffrey; 143 Winged Foot Court, Limerick, PA 19468 (US). DEBOUCK, Christine; 667 Pugh Road, Wayne, PA 19087 (US). ZHONG, YIYI; Apartment D305, 2828 Egypt Road, Audubon, PA 19403 (US). JAWORSKI, Deborah, D.; Apartment 227, 1800 Skiles Boulevard, West Chester, PA 19382 (US). WARREN, Richard, L.; 825 Cathcart Road, Blue Bell, PA 19422 (US). WANG, Min; 302 Centennial Drive, Blue Bell, PA 19422 (US). SHILLING, Lisa, K.; 6 East Park Road, Newtown, PA 18940 (US).  <b>(74) Agent:</b> FISH, Andrew, L.; Dechert Price & Rhoads, 4000 Bell Atlantic Tower, 1717 Arch Street, Philadelphia, PA 19103 (US).  <b>(81) Designated States:</b> JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
<b>(54) Title:</b> NOVEL TIG  <b>(57) Abstract</b>  The invention provides tig polypeptides and polynucleotides encoding tig polypeptides and methods for producing such polypeptides by recombinant techniques. Also provided are methods for utilizing tig polypeptides to screen for antibacterial compounds.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## Novel tig

## RELATED APPLICATIONS

5

## FIELD OF THE INVENTION

10

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, the invention relates to novel polynucleotides and polypeptides of the tig (trigger factor; peptidyl-propyl cis/trans isomerase) family, hereinafter referred to as "tig".

15

## BACKGROUND OF THE INVENTION

The Streptococci make up a medically important genera of microbes known to cause several types of disease in humans, including, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid. Since its isolation more than 100 years ago, *Streptococcus pneumoniae* has been one of the more intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *Streptococcus pneumoniae*, many questions concerning the virulence of this microbe remain. It is particularly preferred to employ Streptococcal genes and gene products as targets for the development of antibiotics.

While certain Streptococcal factors associated with pathogenicity have been identified, e.g., capsule polysaccharides, peptidoglycans, pneumolysins, PspA Complement factor H binding component, autolysin, neuraminidase, peptide permeases, hydrogen peroxide, IgA1 protease, the list is certainly not complete. Moreover, very little is known concerning the temporal expression of such genes during infection and disease progression in a mammalian

host. Discovering the sets of genes the bacterium is likely to be expressing at the different stages of infection, particularly when an infection is established, provides critical information for the screening and characterization of novel antibacterials which can interrupt pathogenesis. In addition to providing a fuller understanding of known proteins, such an approach will  
5 identify previously unrecognised targets.

Trigger factor is a peptidyl-propyl cis/trans isomerase which catalyses proline-limited protein folding extremely well. It has been found associated with nascent protein chains as well as with the chaperone GroEL. The primary physiological defect in *E. coli* trigger factor-depleted or -overproducing cells is an enrichment of filamented cells. Filamentation of the  
10 trigger factor-overproducing strain is suppressed by a multicopy plasmid expressing the essential division gene *ftsZ*, suggesting that trigger factor has an important role in cell division. Therefore, inhibitors of this protein could prevent the bacterium from establishing or maintaining infection of the host and thereby have utility in anti-bacterial therapy.

The frequency of *Streptococcus pneumoniae* infections has risen dramatically in the past  
15 few decades. This has been attributed to the emergence of multiply antibiotic resistant strains and an increasing population of people with weakened immune systems. It is no longer uncommon to isolate *Streptococcus pneumoniae* strains which are resistant to some or all of the standard antibiotics. This phenomenon has created a demand for both new anti-microbial agents, vaccines, and diagnostic tests for this organism.

20 Clearly, there exists a need for factors, such as the tig embodiments of the invention, that have a present benefit of being useful to screen compounds for antibiotic activity. Such factors are also useful to determine their role in pathogenesis of infection, dysfunction and disease. There is also a need for identification and characterization of such factors and their antagonists and agonists to find ways to prevent, ameliorate or correct such infection, dysfunction and disease.

Certain of the polypeptides of the invention possess amino acid sequence homology to a known *B. subtilis* tig protein. (GenBank Accession number: Z75208 Trigger factor (tig) *B. subtilis*; Hesterkamp T. et al., Identification of the prolyl isomerase domain of *Escherichia coli* trigger factor, FEBS Lett 1996 Apr 29;385 (1-2): 67-71; Gothel, S.F. et al., An internal FK506-binding domain is the catalytic core of the prolyl isomerase activity associated with the *Bacillus subtilis* trigger factor, Eur. J. Biochem. 244 (1), 59-65 (1997).)

25

## SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel tig polypeptides by homology between the amino acid sequence set out in Table 1 [SEQ ID NO: 2] and a known amino acid sequence or sequences of other proteins such as *B. subtilis* tig protein.

5 It is a further object of the invention to provide polynucleotides that encode tig polypeptides, particularly polynucleotides that encode the polypeptide herein designated tig.

In a particularly preferred embodiment of the invention, the polynucleotide comprises a region encoding tig polypeptides comprising a sequence set out in Table 1 [SEQ ID NO:1] which includes a full length gene, or a variant thereof.

10 In another particularly preferred embodiment of the invention, there is a novel tig protein from *Streptococcus pneumoniae* comprising the amino acid sequence of Table 1 [SEQ ID NO:2], or a variant thereof.

In accordance with another aspect of the invention, there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in the deposited strain.

15 As a further aspect of the invention, there are provided isolated nucleic acid molecules encoding tig, particularly *Streptococcus pneumoniae* tig, including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

20 In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of tig and polypeptides encoded thereby.

25 As another aspect of the invention, there are provided novel polypeptides of *Streptococcus pneumoniae* referred to herein as tig as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of tig polypeptide encoded by naturally occurring alleles of the tig gene.

30 In a preferred embodiment of the invention, there are provided methods for producing the aforementioned tig polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing tig expression, treating disease, assaying genetic variation, and administering a tig polypeptide or polynucleotide to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

5 In accordance with certain preferred embodiments of this and other aspects of the invention, there are provided polynucleotides that hybridize to tig polynucleotide sequences, particularly under stringent conditions.

In certain preferred embodiments of the invention, there are provided antibodies against tig polypeptides.

10 In other embodiments of the invention, there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or  
15 other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the  
20 compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided tig agonists and antagonists, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention, there are provided compositions comprising a tig polynucleotide or a tig polypeptide for administration to a cell or to a multicellular organism.

25 Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

## DESCRIPTION OF THE INVENTION

30

The invention relates to novel tig polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel tig of

*Streptococcus pneumoniae*, which is related by amino acid sequence homology to *B. subtilis* tig polypeptide. The invention relates especially to tig having the nucleotide and amino acid sequences set out in Table 1 as SEQ ID NO: 1 and SEQ ID NO: 2 respectively, and to the tig nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby.

5

TABLE 1

## tig Polynucleotide and Polypeptide Sequences

(A) Sequences from *Streptococcus pneumoniae* tig polynucleotide sequence [SEQ ID NO:1].

10

5'-1 ATGTCTGTAT CATTTGAAAA CAAAGAAACA AACCGTGGTG TCTTGACTTT

51 CACTATCTCT CAAGACCAAA TCAAACCAGA ATTGGATCGT GTCTTCAAGT

15

101 CAGTGAAGAA ATCTCTTAAT GTTCCAGGTT TCCGTAAAGG TCACCTTCCA

151 CGTCCTATCT TCGATAAAAA ATTTGGTGAA GAATCACTTT ACCAAGACGT

201 TATGAACGCT CTTTGGCCAA ACGCTTATGA AGCAGCTGTA AAAGAAGCTG

20

251 GTCTTGAAGT GGTGCCCCAA CCAAAAATTG ACGTAACTTC AATGGAAAAA

301 GGTCAAGACT GGGTTATCGC TGCTGAAGTC GTTACAAAAC CTGAAGTAAA

25

351 ATTGGGTGAC TACAAAAACC TTGAAGTATC AGTTGATGTA GAAAAAGAAG

401 TAACTGACGC TGATGTTGAA GAGCGTATCG AACGCGAACG CAACAACCTG

451 ACTGAATTGG TTATCAAGGA AGCTGCTGCT GAAAACGGCG ACACTGTTGT

30

501 GATCGACTTC GTTGGTTCTA TCGACGGTGT TGAATTTGAC GGCGGAAAAG

551 GTGAAAACCTT CTCACTTGGA CTTGGTTCAG GTCAATTCAT CCCTGGTTTC

35

601 GAAGACCAAT TGGTAGGTCA CTCAGCTGGC GAAACCGTTG ATGTTATCGT

651 APCATTCCCA GAAGACTACC AAGCAGAGA CCTTGCAGGT AAAGAAGCTA

5  
 701 AATTCGTGAC AACTATCCAT GAAGTAAAAG CTAAAGAAGT TCCAGCTCTT  
 751 GACGATGAAC TTGCAAAGA CATTGATGAA GAAGTTGAAA CACTTGCTGA  
 801 CTTGAAAGAA AAATACCGCA AAGAATTGGC TGCTGCTAAA GAAGAAGCTT  
 851 ACAAAGATGC AGTTGAAGGT GCAGCAATTG ATACAGCTGT AGAAAATGCT  
 10  
 901 GAAATCGTAG AACTTCCAGA AGAAATGATC CACGAAGAAG TTCACCGTTC  
 951 AGTAAATGAA TTCCTTGGGA ACTTGCAACG TCAAGGGATC AACCCTGACA  
 1001 TGTACTTCCA AATCACTGGA ACTACTCAAG AAGACCTTCA CAACCAATAC  
 15  
 1051 CAAGCAGAAG CTGAGTCACG TACTAAGACT AACCTTGTTA TCGAAGCAGT  
 1101 TGCCAAAGCT GAAGGATTG ATGCTTCAGA AGAAGAAATC CAAAAAGAAG  
 20  
 1151 TTGAGCAATT GGCAGCAGAC TACAACATGG AAGTTGCTCA AGTACAAAGC  
 1201 TTGCTTTCAG CTGACATGTT GAAACATGAT ATCACAATCA AAAAAGCTGT  
 1251 TGAATTGATT ACAAGCACAG CAACAGTAAA ATAA -3'

25

(B) *Streptococcus pneumoniae* tig polypeptide sequence deduced from the polynucleotide sequence in this table [SEQ ID NO:2].

NH<sub>2</sub>-1 MSVSFENKET NRGVLTFTIS QDQIKPELDR VFKSVKKSLN VPGFRKGHLP

30

51 RPIFDKKFGE ESLYQDVMNA LLPNAYEAAV KEAGLEVVAQ PKIDVTSMEK

101 GQDWVIAAEV VTKPEVKLGD YKNLEVSVDV EKEVTDADVE ERIERERNL

151 TELVIKEAAA ENGDTVVIDF VGSIDGVEFD GKGGENFSLG LGSGQFIPGF

35

201 EDQLVGHSAG ETVDVIVTFP EDYQAEDLAG KEAKFVTTIH EVKAKEVPAL

251 DDELAKDIDE EVETLADLKE KYRKELAAK EEAYKDAVEG AAIDTAVENA



301 EIVELPEEMI HEEVHRVNE FLGNLQRQGI NPDMYFQITG TTQEDLHNQY

351 QAEAESRTKT NLVIEAVAKA EGFDASEEEI QKEVEQLAAD YNMEVAQVQS

5

401 LLSADMLKHD ITIKKAVELI TSTATVK -COOH

---

#### Deposited materials

10 A deposit containing a *Streptococcus pneumoniae* 0100993 strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (herein "NCIMB"), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned deposit number 40794. The deposit was described as *Streptococcus pneumoniae* 0100993 on deposit. On 17 April 1996 a *Streptococcus pneumoniae* 0100993 DNA library in E. coli was similarly deposited with the  
15 NCIMB and assigned deposit number 40800. The *Streptococcus pneumoniae* strain deposit is referred to herein as "the deposited strain" or as "the DNA of the deposited strain."

The deposited strain contains the full length *tig* gene. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

20 The deposit of the deposited strain has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strain is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required  
25 under 35 U.S.C. §112.

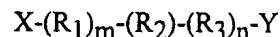
A license may be required to make, use or sell the deposited strain, and compounds derived therefrom, and no such license is hereby granted.

#### Polypeptides

The polypeptides of the invention include a polypeptide of Table 1 [SEQ ID NO:2] (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of *tig*, and also those which have at least 70% identity to a polypeptide of Table 1 [SEQ ID NO:1] or the relevant portion, preferably at least 80% identity to a polypeptide of Table 1 [SEQ ID NO:2] and more preferably at least 90% similarity (more preferably at least 90%  
30

identity) to a polypeptide of Table 1 [SEQ ID NO:2] and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide of Table 1 [SEQ ID NO:2] and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

5 The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal,  $R_1$  and  $R_3$  are any amino acid residue, m is an integer between 1 and 1000 or zero, n is an integer between 1 and 1000 or zero, and  $R_2$  is an amino acid sequence of the invention, particularly  
10 an amino acid sequence selected from Table 1. In the formula above  $R_2$  is oriented so that its amino terminal residue is at the left, bound to  $R_1$ , and its carboxy terminal residue is at the right, bound to  $R_3$ . Any stretch of amino acid residues denoted by either R group, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same  
15 as part but not all of the amino acid sequence of the aforementioned polypeptides. As with tig polypeptides fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence of Table 1 [SEQ ID NO:2], or of variants thereof, such as a continuous series  
20 of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and  
25 coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of tig, including those with a similar activity or an improved activity, or with a decreased  
30 undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains

of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants  
5 may be employed as intermediates for producing the full-length polypeptides of the invention.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" may also be used in describing certain polypeptides of the invention. "X" and "Xaa" mean that any of the twenty naturally occurring amino acids may appear at such a designated position in the polypeptide sequence.

#### 10 Polynucleotides

Another aspect of the invention relates to isolated polynucleotides, including the full length gene, that encode the tig polypeptide having a deduced amino acid sequence of Table 1 [SEQ ID NO:2] and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as a polynucleotide sequence set out in Table 1  
15 [SEQ ID NO:1], a polynucleotide of the invention encoding tig polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence given in Table 1 [SEQ ID NO:1], typically a library of clones of  
20 chromosomal DNA of *Streptococcus pneumoniae* 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the  
25 sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and  
30 Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotide set out in Table 1 [SEQ ID NO:1] was discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

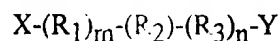
The DNA sequence set out in Table 1 [SEQ ID NO:1] contains an open reading frame encoding a protein having about the number of amino acid residues set forth in Table 1 [SEQ ID NO:2] with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known in the art. The polynucleotide of SEQ ID NO: 1, between nucleotide  
 5 number 1 and the stop codon which begins at nucleotide number 1282 of SEQ ID NO:1, encodes the polypeptide of SEQ ID NO:2.

The tig polypeptide of the invention is structurally related to other proteins of the tig (peptidyl-propyl cis/trans isomerase) family, as shown by the results of sequencing the DNA encoding tig of the deposited strain.

10 The invention provides a polynucleotide sequence identical over its entire length to a coding sequence in Table 1 [SEQ ID NO:1]. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may  
 15 also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the  
 20 marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

25 A preferred embodiment of the invention is a polynucleotide of comprising nucleotide 1 to the nucleotide immediately upstream of or including nucleotide 1282 set forth in SEQ ID NO:1 of Table 1, both of which encode the tig polypeptide.

The invention also includes polynucleotides of the formula:



30 wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal,  $R_1$  and  $R_3$  is any nucleic acid residue,  $m$  is an integer between 1 and 3000 or zero,  $n$  is an integer between 1 and 3000 or zero, and  $R_2$  is a nucleic acid sequence of the invention,

particularly a nucleic acid sequence selected from Table 1. In the polynucleotide formula above  $R_2$  is oriented so that its 5' end residue is at the left, bound to  $R_1$ , and its 3' end residue is at the right, bound to  $R_3$ . Any stretch of nucleic acid residues denoted by either R group, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In a preferred embodiment m and/or n is an integer between 1 and 1000.

It is most preferred that the polynucleotides of the inventions are derived from *Streptococcus pneumoniae*, however, they may preferably be obtained from organisms of the same taxonomic genus. They may also be obtained, for example, from organisms of the same taxonomic family or order.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* tig having an amino acid sequence set out in Table 1 [SEQ ID NO:2]. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having a deduced amino acid sequence of Table 1 [SEQ ID NO:2]. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding tig variants, that have the amino acid sequence of tig polypeptide of Table 1 [SEQ ID NO:2] in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of tig.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to a polynucleotide encoding tig polypeptide having an amino acid sequence set out in Table 1 [SEQ ID NO:2], and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding tig polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these

particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

5 Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by a DNA of Table 1 [SEQ ID NO:1].

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under  
10 stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate  
15 (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

20 The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NO:1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NO:1 or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a  
25 polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding *tig* and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the *tig*  
30 gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the *tig* gene may be isolated by screening using a DNA sequence provided in Table 1 [SEQ ID NO: 1] to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the sequences of Table 1 [SEQ ID NOS:1 or 2] may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" may also be used in describing certain polynucleotides of the invention. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein

having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

5           **Vectors, host cells, expression**

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

10           For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring  
15 Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci,  
20 staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the  
25 invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those  
30 derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express



polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

5 For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures  
10 by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active  
15 conformation when the polypeptide is denatured during isolation and or purification.

#### Diagnostic Assays

This invention is also related to the use of the tig polynucleotides of the invention for use as diagnostic reagents. Detection of tig in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"),  
20 particularly mammals, and especially humans, particularly those infected or suspected to be infected with an organism comprising the tig gene may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection  
25 or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA, cDNA and genomic DNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point  
30 mutations can be identified by hybridizing amplified DNA to labeled tig polynucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by

alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, e.g., Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985).

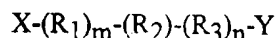
Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA, cDNA or genomic DNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding tig can be used to identify and analyze mutations. Examples of representative primers are shown below in Table 2.

Table 2

Primers for amplification of tig polynucleotides

<u>SEQ ID NO</u>	<u>PRIMER SEQUENCE</u>
3	5'-ATTCTTTGCGGTATTTTCTTTCA-3'
4	5'-GTTGGTTCTATCGACGGTGTG-3'

The invention also includes primers of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal,  $R_1$  and  $R_3$  is any nucleic acid residue, m is an integer between 1 and 20 or zero, n is an integer between 1 and 20 or zero, and  $R_2$  is a primer sequence of the invention, particularly a primer sequence selected from Table 2. In the polynucleotide formula above  $R_2$  is oriented so that its 5' end residue is at the left, bound to  $R_1$ , and its 3' end residue is at the right, bound to  $R_3$ . Any stretch of nucleic acid residues denoted by either R group, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer being complementary to a region of a polynucleotide of Table 1. In a preferred embodiment m and/or n is an integer between 1 and 10.

The invention further provides these primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. These primers may be used for, among other things, amplifying tig DNA

isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

5       The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having a sequence of Table 1 [SEQ ID NO: 1]. Increased or decreased expression of tig polynucleotide can be measured using any on of the methods well known in the art for the quantation of  
10 polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of tig protein compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a tig  
15 protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

#### Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used  
20 as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by  
25 administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in  
30 *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with  
5 binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-tig or from naive libraries (McCafferty, J. et al., (1990), *Nature* 348, 552-554; Marks, J. et al., (1992) *Biotechnology* 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) *Nature* 352, 624-628).

10 If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against tig- polypeptide may be employed to treat infections,  
15 particularly bacterial infections.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide  
20 according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

25 The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple  
30 antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described  
5 in Jones, P. et al. (1986), *Nature* 321, 522-525 or Tempest et al., (1991) *Biotechnology* 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., *Hum Mol Genet* 1992, 1:363, Manthorpe et al., *Hum. Gene Ther.* 1993:4, 419), delivery  
10 of DNA complexed with specific protein carriers (Wu et al., *J Biol Chem.* 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS USA*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., *Science* 1989:243,375), particle bombardment (Tang et al., *Nature* 1992, 356:152, Eisenbraun et al., *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., *PNAS USA* 1984:81,5849).  
15

#### Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be  
20 structural or functional mimetics. See, e.g., Coligan et al., *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of tig polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve  
25 high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising tig polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a tig agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the tig polypeptide is reflected in decreased  
30 binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, i.e., without inducing the effects of tig polypeptide are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are

agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in tig polynucleotide or polypeptide activity, and binding assays known in the art.

5 Another example of an assay for tig antagonists is a competitive assay that combines tig and a potential antagonist with tig-binding molecules, recombinant tig binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. The tig molecule can be labeled, such as by radioactivity or a colorimetric compound, such that the number of tig molecules bound to a binding molecule or converted to  
10 product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a  
15 binding molecule, without inducing tig-induced activities, thereby preventing the action of tig by excluding tig from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic  
20 molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of tig.

25 Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to  
30 control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and

mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block tig protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial tig proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat diseases.

*Helicobacter pylori* (herein *H. pylori*) bacteria infect the stomachs of over one-third of the world's population causing stomach cancer, ulcers, and gastritis (International Agency for Research on Cancer (1994) Schistosomes, Liver Flukes and *Helicobacter Pylori* (International Agency for Research on Cancer, Lyon, France; <http://www.uicc.ch/ecp/ecp2904.htm>). Moreover, the international Agency for Research on Cancer recently recognized a cause-and-effect relationship between *H. pylori* and gastric adenocarcinoma, classifying the bacterium as a Group I (definite) carcinogen. Preferred antimicrobial compounds of the invention (agonists and antagonists of tig) found using screens provided by the invention, particularly broad-spectrum antibiotics, should be useful in the treatment of *H. pylori* infection. Such treatment should decrease the advent of *H. pylori*-induced cancers, such as gastrointestinal carcinoma. Such treatment should also cure gastric ulcers and gastritis.

#### Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with tig, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of tig, or a fragment or a variant thereof, for expressing tig, or a fragment or a variant thereof *in vivo* in order to induce an

immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a tig or protein coded therefrom, wherein the composition comprises a recombinant tig or protein coded therefrom comprising DNA which codes for and expresses an antigen of said tig or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A tig polypeptide or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of



monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

5       The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary  
10       glands, urethra or vagina.

      The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations  
15       suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example,  
20       sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

25       While the invention has been described with reference to certain tig protein, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

#### Compositions, kits and administration

30       The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms,

such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof.

- 5 The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

- 10 The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

- 15 Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also  
20 contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

- For administration to mammals, and particularly humans, it is expected that the daily  
25 dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of  
30 this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended

time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

5 The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

10 Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

15 In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

20 Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1 µg/ml to 10mg/ml for bathing of wounds or indwelling devices.

25 A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

30

## GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Disease(s)" means and disease caused by or related to infection by a bacteria, including otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences.

10 In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Methods to determine identity are designed to give the

20 largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990). The BLAST X program is publicly available

25 from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following: Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

30 Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci. USA.* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

- 5           Parameters for polynucleotide comparison include the following: Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

- 10          Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

- (1)       Polynucleotide embodiments further include an isolated polynucleotide
- 15       comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to the reference sequence of SEQ ID NO:1, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO:1 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including
- 20       transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID
- 25       NO:1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

- 30       wherein  $n_n$  is the number of nucleotide alterations,  $x_n$  is the total number of nucleotides in SEQ ID NO:1,  $y$  is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and  $\bullet$  is the symbol for the multiplication

operator, and wherein any non-integer product of  $x_n$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_n$ . Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is it may be 100% identical, or it may include up to a certain integer number of nucleic acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of nucleic acids in SEQ ID NO:1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleic acids in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein  $n_n$  is the number of nucleic acid alterations,  $x_n$  is the total number of nucleic acids in SEQ ID NO:1,  $y$  is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc.,  $\bullet$  is the symbol for the multiplication operator, and wherein any non-integer product of  $x_n$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_n$ .

(2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to a polypeptide reference sequence of SEQ ID NO:2, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO:2 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those

terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting  
 5 that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein  $n_a$  is the number of amino acid alterations,  $x_a$  is the total number of amino acids in SEQ  
 10 ID NO:2,  $y$  is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and  $\bullet$  is the symbol for the multiplication operator, and wherein any non-integer product of  $x_a$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

By way of example, a polypeptide sequence of the present invention may be identical to  
 15 the reference sequence of SEQ ID NO:2, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or  
 20 carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then  
 25 subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein  $n_a$  is the number of amino acid alterations,  $x_a$  is the total number of amino acids in SEQ  
 30 ID NO:2,  $y$  is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and  $\bullet$  is the symbol

for the multiplication operator, and wherein any non-integer product of  $x_a$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxiribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally



referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide

changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are  
5 closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-  
10 naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

## EXAMPLES

15 The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

### Example 1 Strain selection, Library Production and Sequencing

The polynucleotide having a DNA sequence given in Table 1 [SEQ ID NO:1] was  
20 obtained from a library of clones of chromosomal DNA of *Streptococcus pneumoniae* in *E. coli*. The sequencing data from two or more clones containing overlapping *Streptococcus pneumoniae* DNAs was used to construct the contiguous DNA sequence in SEQ ID NO:1. Libraries may be prepared by routine methods, for example:

Methods 1 and 2 below.

25 Total cellular DNA is isolated from *Streptococcus pneumoniae* 0100993 according to standard procedures and size-fractionated by either of two methods.

#### Method 1

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are  
30 rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library

packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

#### Method 2

Total cellular DNA is partially hydrolyzed with a one or a combination of restriction enzymes appropriate to generate a series of fragments for cloning into library vectors (e.g., RsaI, PstI, AluI, BshI235I), and such fragments are size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

#### Example 2

**The determination of expression during infection of a gene from *Streptococcus pneumoniae***

Excised lungs from a 48 hour respiratory tract infection of *Streptococcus pneumoniae* 0100993 in the mouse is efficiently disrupted and processed in the presence of chaotropic agents and RNAase inhibitor to provide a mixture of animal and bacterial RNA. The optimal conditions for disruption and processing to give stable preparations and high yields of bacterial RNA are followed by the use of hybridisation to a radiolabelled oligonucleotide specific to *Streptococcus pneumoniae* 16S RNA on Northern blots. The RNAase free, DNAase free, DNA and protein free preparations of RNA obtained are suitable for Reverse Transcription PCR (RT-PCR) using unique primer pairs designed from the sequence of each gene of *Streptococcus pneumoniae* 0100993.

#### a) Isolation of tissue infected with *Streptococcus pneumoniae* 0100993 from a mouse animal model of infection (lungs)

*Streptococcus pneumoniae* 0100993 is grown either on TSA/5%horse blood plates or in AGCH medium overnight, 37°C, 5%CO<sub>2</sub>. Bacteria are then collected and resuspended in phosphate-buffered saline to an A<sub>600</sub> of approximately 0.4. Mice are anaesthetized with isofluorane and 50ml of bacterial suspension (approximately 2 x 10<sup>5</sup> bacteria) is administered intranasally using a pipetman. Mice are allowed to recover and have food and water ad libitum. After 48 hours, the mice are euthanized by carbon dioxide overdose, and lungs are aseptically removed and snap-frozen in liquid nitrogen.

b) Isolation of *Streptococcus pneumoniae* 0100993 RNA from infected tissue samples

Infected tissue samples, in 2-ml cryo-storage tubes, are removed from -80°C storage into a dry ice ethanol bath. In a microbiological safety cabinet the samples are disrupted up to  
5 eight at a time while the remaining samples are kept frozen in the dry ice ethanol bath. To  
disrupt the bacteria within the tissue sample, 50-100 mg of the tissue is transferred to a  
FastRNA tube containing a silica/ceramic matrix (BIO101). Immediately, 1 ml of extraction  
reagents (FastRNA reagents, BIO101) are added to give a sample to reagent volume ratio of  
approximately 1 to 20. The tubes are shaken in a reciprocating shaker (FastPrep FP120,  
10 BIO101) at 6000 rpm for 20-120 sec. The crude RNA preparation is extracted with  
chloroform/isoamyl alcohol, and precipitated with DEPC-treated/Isopropanol Precipitation  
Solution (BIO101). RNA preparations are stored in this isopropanol solution at -80°C if  
necessary. The RNA is pelleted (12,000g for 10 min.), washed with 75% ethanol (v/v in  
DEPC-treated water), air-dried for 5-10 min, and resuspended in 0.1 ml of DEPC-treated  
15 water, followed by 5-10 minutes at 55 °C. Finally, after at least 1 minute on ice, 200 units of  
Rnasin (Promega) is added.

RNA preparations are stored at -80°C for up to one month. For longer term storage the  
RNA precipitate can be stored at the wash stage of the protocol in 75% ethanol for at least one  
year at -20°C.

20 Quality of the RNA isolated is assessed by running samples on 1% agarose gels. 1 x  
TBE gels stained with ethidium bromide are used to visualise total RNA yields. To demonstrate  
the isolation of bacterial RNA from the infected tissue 1 x MOPS, 2.2M formaldehyde gels are  
run and vacuum blotted to Hybond-N (Amersham). The blot is then hybridised with a <sup>32</sup>P-  
labelled oligonucleotide probe, of sequence 5'- AACTGAGACTGGCTTTAAGAGATTA -3' [SEQ  
25 ID NO:5], specific to 16S rRNA of *Streptococcus pneumoniae*. The size of the hybridising  
band is compared to that of control RNA isolated from *in vitro* grown *Streptococcus  
pneumoniae* 0100993 in the Northern blot. Correct sized bacterial 16S rRNA bands can be  
detected in total RNA samples which show degradation of the mammalian RNA when  
visualised on TBE gels.

30

c) The removal of DNA from *Streptococcus pneumoniae*-derived RNA

DNA was removed from 50 microgram samples of RNA by a 30 minute treatment at 37°C with 20 units of RNAase-free DNAaseI (GenHunter) in the buffer supplied in a final volume of 57 microliters.

5 The DNAase was inactivated and removed by treatment with TRIzol LS Reagent (Gibco BRL, Life Technologies) according to the manufacturers protocol.

DNAase treated RNA was resuspended in 100 microlitres of DEPC treated water with the addition of Rnasin as described before.

d) The preparation of cDNA from RNA samples derived from infected tissue

10 3 microgram samples of DNAase treated RNA are reverse transcribed using a SuperScript Preamplification System for First Strand cDNA Synthesis kit (Gibco BRL, Life Technologies) according to the manufacturers instructions. 150 nanogram of random hexamers is used to prime each reaction. Controls without the addition of SuperScriptII reverse transcriptase are also run. Both +/-RT samples are treated with RNaseH before proceeding to  
15 the PCR reaction

e) The use of PCR to determine the presence of a bacterial cDNA species

PCR reactions are set up on ice in 0.2ml tubes by adding the following components: 43 microlitres PCR Master Mix (Advanced Biotechnologies Ltd.); 1 microlitre PCR primers  
20 (optimally 18-25 basepairs in length and designed to possess similar annealing temperatures), each primer at 10mM initial concentration; and 5 microlitres cDNA.

PCR reactions are run on a Perkin Elmer GeneAmp PCR System 9600 as follows: 2 minutes at 94 °C, then 50 cycles of 30 seconds each at 94 °C, 50 °C and 72 °C followed by 7 minutes at 72 °C and then a hold temperature of 20 °C. (the number of cycles is optimally 30-50  
25 to determine the appearance or lack of a PCR product and optimally 8-30 cycles if an estimation of the starting quantity of cDNA from the RT reaction is to be made); 10 microlitre aliquots are then run out on 1% 1 x TBE gels stained with ethidium bromide, with PCR product, if present, sizes estimated by comparison to a 100 bp DNA Ladder (Gibco BRL, Life Technologies). Alternatively if the PCR products are conveniently labelled by the use of a  
30 labelled PCR primer (e.g. labelled at the 5'end with a dye) a suitable aliquot of the PCR product is run out on a polyacrylamide sequencing gel and its presence and quantity detected using a suitable gel scanning system (e.g. ABI Prism™ 377 Sequencer using GeneScan™ software as supplied by Perkin Elmer).

RT/PCR controls may include +/- reverse transcriptase reactions, 16S rRNA primers or DNA specific primer pairs designed to produce PCR products from non-transcribed *Streptococcus pneumoniae* 0100993 genomic sequences.

To test the efficiency of the primer pairs they are used in DNA PCR with  
5 *Streptococcus pneumoniae* 0100993 total DNA. PCR reactions are set up and run as described above using approx. 1 microgram of DNA in place of the cDNA.

Primer pairs which fail to give the predicted sized product in either DNA PCR or RT/PCR are PCR failures and as such are uninformative. Of those which give the correct size product with DNA PCR two classes are distinguished in RT/PCR: 1. Genes which are not  
10 transcribed *in vivo* reproducibly fail to give a product in RT/PCR; and 2. Genes which are transcribed *in vivo* reproducibly give the correct size product in RT/PCR and show a stronger signal in the +RT samples than the signal (if at all present) in -RT controls.

**What is claimed is:**

1. An isolated polynucleotide comprising a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2.
2. An isolated polynucleotide comprising a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the *tig* gene contained in the *Streptococcus pneumoniae* of the deposited strain.
3. An isolated polynucleotide comprising a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2.
4. An isolated polynucleotide that is complementary to the polynucleotide of claim 1
5. The polynucleotide of Claim 1 wherein the polynucleotide is DNA or RNA
6. The polynucleotide of Claim 1 comprising the nucleic acid sequence set forth in SEQ ID NO:1.
7. The polynucleotide of Claim 1 comprising nucleotide 1 to the stop codon which begins at nucleotide number 1282 set forth in SEQ ID NO:1.
8. The polynucleotide of Claim 1 which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.
9. A vector comprising the polynucleotide of Claim 1.
10. A host cell comprising the vector of Claim 9.
11. A process for producing a polypeptide comprising: expressing from the host cell of Claim 10 a polypeptide encoded by said DNA.
12. A process for producing a *tig* polypeptide or fragment comprising culturing a host of claim 10 under conditions sufficient for the production of said polypeptide or fragment.
13. A polypeptide comprising an amino acid sequence which is at least 70% identical to the amino acid sequence of SEQ ID NO:2.
14. A polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2.
15. An antibody against the polypeptide of claim 14.
16. An antagonist which inhibits the activity or expression of the polypeptide of claim 14.
17. A method for the treatment of an individual in need of *tig* polypeptide comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 14.

18. A method for the treatment of an individual having need to inhibit tig polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 14.

19. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 14 in an individual comprising:

- (a) determining a nucleic acid sequence encoding said polypeptide, and/or
- (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.

20. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 14 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;

and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.

21. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with tig polypeptide of claim 14, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.

22. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of tig polypeptide of claim 14, or fragment or a variant thereof, for expressing said tig polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.



<110> SmithKline Beecham Corporation

<130> GM10118

<151> 1997-11-20

<170> FastSEQ for Windows Version 3.0

<211> 1284

<212> DNA

<213> Streptococcus pneumoniae

<220>

<221> CDS

 $\langle 222 \rangle \quad (1) \dots (1281)$ 

<400> 1

atg tct gta tca ttt gaa aac aaa gaa aca aac cgt ggt gtc ttg act 48  
Met Ser Val Ser Phe Glu Asn Lys Glu Thr Asn Arg Gly Val Leu Thr  
1 5 10 15

ttc act atc tct caa gac caa atc aaa cca gaa ttg gat cgt gtc ttc 96  
Phe Thr Ile Ser Gln Asp Gln Ile Lys Pro Glu Leu Asp Arg Val Phe  
20 25 30

aag tca gtg aag aaa tct ctt aat gtt cca ggt ttc cgt aaa ggt cac 144  
Lys Ser Val Lys Lys Ser Leu Asn Val Pro Gly Phe Arg Lys Gly His  
35 40 45

ctt cca cgt cct atc ttc gat aaa aaa ttt ggt gaa gaa tca ctt tac 192  
Leu Pro Arg Pro Ile Phe Asp Lys Lys Phe Gly Glu Glu Ser Leu Tyr

50	55	60	
caa gac gtt atg aac gct ctt ttg cca aac gct tat gaa gca gct gta			240
Gln Asp Val Met Asn Ala Leu Leu Pro Asn Ala Tyr Glu Ala Ala Val			
65	70	75	80
aaa gaa gct ggt ctt gaa gtg gtt gcc caa cca aaa att gac gta act			288
Lys Glu Ala Gly Leu Glu Val Val Ala Gln Pro Lys Ile Asp Val Thr			
85	90	95	
tca atg gaa aaa ggt caa gac tgg gtt atc gct gct gaa gtc gtt aca			336
Ser Met Glu Lys Gly Gln Asp Trp Val Ile Ala Ala Glu Val Val Thr			
100	105	110	
aaa cct gaa gta aaa ttg ggt gac tac aaa aac ctt gaa gta tca gtt			384
Lys Pro Glu Val Lys Leu Gly Asp Tyr Lys Asn Leu Glu Val Ser Val			
115	120	125	
gat gta gaa aaa gaa gta act gac gct gat gtt gaa gag cgt atc gaa			432
Asp Val Glu Lys Glu Val Thr Asp Ala Asp Val Glu Glu Arg Ile Glu			
130	135	140	
cgc gaa cgc aac aac ctg act gaa ttg gtt atc aag gaa gct gct gct			480
Arg Glu Arg Asn Asn Leu Thr Glu Leu Val Ile Lys Glu Ala Ala Ala			
145	150	155	160
gaa aac ggc gac act gtt gtg atc gac ttc gtt ggt tct atc gac ggt			528
Glu Asn Gly Asp Thr Val Val Ile Asp Phe Val Gly Ser Ile Asp Gly			
165	170	175	
gtt gaa ttt gac ggc gga aaa ggt gaa aac ttc tca ctt gga ctt ggt			576
Val Glu Phe Asp Gly Gly Lys Gly Glu Asn Phe Ser Leu Gly Leu Gly			
180	185	190	
tca ggt caa ttc atc cct ggt ttc gaa gac caa ttg gta ggt cac tca			624
Ser Gly Gln Phe Ile Pro Gly Phe Glu Asp Gln Leu Val Gly His Ser			
195	200	205	
gct ggc gaa acc gtt gat gtt atc gta aca ttc cca gaa gac tac caa			672
Ala Gly Glu Thr Val Asp Val Ile Val Thr Phe Pro Glu Asp Tyr Gln			
210	215	220	

aaa gct gaa gga ttt gat gct tca gaa gaa gaa atc caa aaa gaa gtt 1152  
Lys Ala Glu Gly Phe Asp Ala Ser Glu Glu Glu Ile Gln Lys Glu Val  
370 375 380

gag caa ttg gca gca gac tac aac atg gaa gtt gct caa gta caa agc 1200  
 Glu Gln Leu Ala Ala Asp Tyr Asn Met Glu Val Ala Gln Val Gln Ser  
 385 390 395 400

ttg ctt tca gct gac atg ttg aaa cat gat atc aca atc aaa aaa gct 1248  
 Leu Leu Ser Ala Asp Met Leu Lys His Asp Ile Thr Ile Lys Lys Ala  
 405 410 415

gtt gaa ttg att aca agc aca gca aca gta aaa taa 1284  
 Val Glu Leu Ile Thr Ser Thr Ala Thr Val Lys  
 420 425

<210> 2

<211> 427

<212> PRT

<213> Streptococcus pneumoniae

<400> 2

Met Ser Val Ser Phe Glu Asn Lys Glu Thr Asn Arg Gly Val Leu Thr  
 1 5 10 15  
 Phe Thr Ile Ser Gln Asp Gln Ile Lys Pro Glu Leu Asp Arg Val Phe  
 20 25 30  
 Lys Ser Val Lys Lys Ser Leu Asn Val Pro Gly Phe Arg Lys Gly His  
 35 40 45  
 Leu Pro Arg Pro Ile Phe Asp Lys Lys Phe Gly Glu Glu Ser Leu Tyr  
 50 55 60  
 Gln Asp Val Met Asn Ala Leu Leu Pro Asn Ala Tyr Glu Ala Ala Val  
 65 70 75 80  
 Lys Glu Ala Gly Leu Glu Val Val Ala Gln Pro Lys Ile Asp Val Thr  
 85 90 95  
 Ser Met Glu Lys Gly Gln Asp Trp Val Ile Ala Ala Glu Val Val Thr  
 100 105 110  
 Lys Pro Glu Val Lys Leu Gly Asp Tyr Lys Asn Leu Glu Val Ser Val  
 115 120 125  
 Asp Val Glu Lys Glu Val Thr Asp Ala Asp Val Glu Glu Arg Ile Glu  
 130 135 140  
 Arg Glu Arg Asn Asn Leu Thr Glu Leu Val Ile Lys Glu Ala Ala Ala  
 145 150 155 160  
 Glu Asn Gly Asp Thr Val Val Ile Asp Phe Val Gly Ser Ile Asp Gly  
 165 170 175

Val Glu Phe Asp Gly Gly Lys Gly Glu Asn Phe Ser Leu Gly Leu Gly  
 180 185 190  
 Ser Gly Gln Phe Ile Pro Gly Phe Glu Asp Gln Leu Val Gly His Ser  
 195 200 205  
 Ala Gly Glu Thr Val Asp Val Ile Val Thr Phe Pro Glu Asp Tyr Gln  
 210 215 220  
 Ala Glu Asp Leu Ala Gly Lys Glu Ala Lys Phe Val Thr Thr Ile His  
 225 230 235 240  
 Glu Val Lys Ala Lys Glu Val Pro Ala Leu Asp Asp Glu Leu Ala Lys  
 245 250 255  
 Asp Ile Asp Glu Glu Val Glu Thr Leu Ala Asp Leu Lys Glu Lys Tyr  
 260 265 270  
 Arg Lys Glu Leu Ala Ala Ala Lys Glu Glu Ala Tyr Lys Asp Ala Val  
 275 280 285  
 Glu Gly Ala Ala Ile Asp Thr Ala Val Glu Asn Ala Glu Ile Val Glu  
 290 295 300  
 Leu Pro Glu Glu Met Ile His Glu Glu Val His Arg Ser Val Asn Glu  
 305 310 315 320  
 Phe Leu Gly Asn Leu Gln Arg Gln Gly Ile Asn Pro Asp Met Tyr Phe  
 325 330 335  
 Gln Ile Thr Gly Thr Thr Gln Glu Asp Leu His Asn Gln Tyr Gln Ala  
 340 345 350  
 Glu Ala Glu Ser Arg Thr Lys Thr Asn Leu Val Ile Glu Ala Val Ala  
 355 360 365  
 Lys Ala Glu Gly Phe Asp Ala Ser Glu Glu Glu Ile Gln Lys Glu Val  
 370 375 380  
 Glu Gln Leu Ala Ala Asp Tyr Asn Met Glu Val Ala Gln Val Gln Ser  
 385 390 395 400  
 Leu Leu Ser Ala Asp Met Leu Lys His Asp Ile Thr Ile Lys Lys Ala  
 405 410 415  
 Val Glu Leu Ile Thr Ser Thr Ala Thr Val Lys  
 420 425

&lt;210&gt; 3

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Streptococcus pneumoniae

&lt;400&gt; 3

attctttgcg gtatttttct ttca

<210> 4

<211> 22

<212> DNA

<213> Streptococcus pneumoniae

<400> 4

g ttg gtt cta t c g a c g g t g t t g

22

<210> 5

<211> 25

<212> DNA

<213> Streptococcus pneumoniae

<400> 5

a a c t g a g a c t g g c t t t a a g a g a t t a

25

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/24805

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 1/21, 9/90, 15/61, 15/63  
US CL : 435/233, 320.1, 252.3; 536/23.1, 23.2

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/233, 320.1, 252.3; 536/23.1, 23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
SEQ ID Nos: 1-2 searched on all commercial databases

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS and DIALOG (Biotech)

search terms: tig, trigger factor, peptidyl-propyl cis/trans isomerase, streptococcus

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GOTHEL et al. An internal FK506-binding domain is the catalytic core of the prolyl isomerase activity associated with the <i>Bacillus subtilis</i> trigger factor. Eur. J. Biochem. October 1997, Vol. 244, pages 59-65.	1-22

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"A" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

09 MARCH 1999

Date of mailing of the international search report

23 MAR 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRADLEY S. MAYHEW

Telephone No. (703) 308-0196